

CLAIMS

What is claimed is:

1. A computer-implemented method comprising:
parsing a plurality of documents;
selecting paragraphs from the documents through subsuming relation calculation;
and
rewriting the selected paragraphs into a summary.
2. The method of claim 1 wherein parsing further comprises:
extracting noun phrases and verb phrases from the documents;
categorizing the noun phrases that are entity names; and
converting the entity names into canonical form.
3. The method of claim 1 wherein subsuming relation calculation further comprises:
linking noun phrases, verb phrases or entity names in each paragraph of every
document with identical noun phrases, verb phrases or entity names in every other
paragraph of every document;
and
counting the links for each paragraph.
4. The method of claim 1 wherein rewriting further comprises:
ranking the paragraphs;

applying a co-reference resolution algorithm to the paragraphs; and
replacing pronouns in the paragraphs with their full entity name antecedents.

5. The method of claim 1 wherein the documents have a common topic independent of domain.

6. The method of claim 1 wherein the documents are composed in English or in a language other than English.

7. A machine readable medium having stored thereon sequences of instructions which are executable by a processor, and which, when executed by the processor, cause the system to perform a method comprising:

parsing a plurality of documents;

selecting paragraphs from the documents through subsuming relation calculation;

and

rewriting the selected paragraphs into a summary.

8. The medium of claim 7 wherein parsing further comprises:
extracting noun phrases and verb phrases from the documents;
categorizing the noun phrases that are entity names; and
converting the entity names into canonical form.

9. The medium of claim 7 wherein subsuming relation calculation further comprises:

linking noun phrases, verb phrases or entity names in each paragraph of every document with identical noun phrases, verb phrases or entity names in every other paragraph of every document; and
counting the links for each paragraph.

10. The medium of claim 7 wherein rewriting further comprises:
ranking the paragraphs;
applying a co-reference resolution algorithm to the paragraphs; and
replacing pronouns in the paragraphs with their full entity name antecedents.
11. The medium of claim 7 wherein the documents have a common topic independent of domain.
12. The medium of claim 7 wherein the documents are composed in English or in a language other than English.
13. An system comprising:
a processor;
a bus coupled to the processor; and
a unit coupled to the bus to parse a plurality of documents, select paragraphs from the documents through subsuming relation calculation, and rewrite the selected paragraphs into a summary.

14. The system of claim 13 wherein the unit further extracts noun phrases and verb phrases from the documents, categorizes the noun phrases that are entity names, and converts the entity names into canonical form.
15. The system of claim 13 wherein the unit further links noun phrases, verb phrases or entity names in each paragraph of every document with identical noun phrases, verb phrases or entity names in every other paragraph of every document, and counts the links for each paragraph.
16. The system of claim 13 wherein the unit further ranks the paragraphs, applies a co-reference resolution algorithm to the paragraphs, and replaces pronouns in the paragraphs with their full entity name antecedents.
17. The system of claim 13 wherein the documents have a common topic independent of domain.
18. The system of claim 13 wherein the documents are composed in English or in a language other than English.